

Wnioskowanie bayesowskie – podstawowe pojęcia

1. Podstawowe pojęcia w analizie bayesowskiej

Obiekt	Wyjaśnienie	Uwagi
θ	<p>Wektor (kolumnowy) <i>wszystkich</i> parametrów modelu – zarówno tych interesujących nas (ang. <i>parameters of interest</i>), jak i parametrów ubocznych, zakłócających (ang. <i>nuisance parameters</i>). Wektor θ stanowi element <u>przestrzeni parametrów</u>, Θ, tj. $\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_s]' \in \Theta \subseteq \mathbb{R}^s$ (gdzie \mathbb{R} jest zbiorem liczb rzeczywistych, zaś s oznacza wymiar przestrzeni Θ, tzn. $s = \dim\Theta$). Z reguły mamy</p> $\Theta = \bigotimes_{j=1}^s \Theta_j,$ <p>tzn. przestrzeń parametrów jest iloczynem kartezjańskim przestrzeni wszystkich pojedynczych parametrów, $\Theta_j \subseteq \mathbb{R}$, dla $j = 1, 2, \dots, s$.</p>	W odróżnieniu od modeli rozważanych na gruncie podejścia częstościowego, w modelu bayesowskim parametry traktowane są jako <i>zmiennie losowe</i> .
y	Wektor obserwacji (tych dostępnych, a nie np. prognozowanych) – stanowi element zbioru Y , tj. $y \in Y \subseteq \mathbb{R}^T$, gdzie T oznacza liczbę dostępnych obserwacji.	Jeśli nie zajmujemy się modelowaniem obserwacji niedostępnych (w tym prognozowaniem), to Y jest <i>przestrzenią obserwacji</i> . W przeciwnym razie, przestrzenią obserwacji (dostępnych i niedostępnych) nazywamy zbiór $Y \times Y^f \subseteq \mathbb{R}^{T+h}$.
y^f	Wektor obserwacji niedostępnych – zwykle stanowią go wartości prognozowane lub „niezaobserwowane obserwacje”, czyli te, dla których brakuje danych (z tym ostatnim możemy mieć do czynienia np. w przypadku finansowych szeregów czasowych o dziennej częstotliwości – gdy giełda jest nieczynna w danym dniu, notowania zostają zawieszane). Wektor y^f stanowi element zbioru Y^f , tj. $y^f \in Y^f \subseteq \mathbb{R}^h$, gdzie h oznacza liczbę obserwacji niedostępnych.	
UWAGA: W dalszym ciągu ograniczamy się (tymczasowo) do modelowania tylko obserwacji dostępnych; prognozowaniem zajmiemy się dopiero za jakiś czas.		
$p(\cdot)$	Ogólne oznaczenie funkcji gęstości pewnej zmiennej losowej (wektora losowego). Argument gęstości określa zmienną losową, której ta gęstość dotyczy. Dopuszczając się pewnego nadużycia notacji, przyjmujemy konwencję, w której symbol $p(x)$ odnosi się także do samego <i>rozkładu</i> zmiennej losowej x . ¹ Cała ta konwencja jest powszechnie stosowana w literaturze bayesowskiej. Z drugiej strony, specjalną notacją stosujemy dla gęstości znanych rozkładów prawdopodobieństwa, np. $f_N(\cdot)$ oznacza funkcję gęstości rozkładu normalnego.	
$p(y, \theta)$	Model bayesowski – utożsamiany z łącznym rozkładem obserwacji i parametrów. (Nieco ściślej $p(y, \theta)$ oznacza <i>funkcję gęstości</i> łącznego rozkładu wektora obserwacji i parametrów – patrz przypis nr 1).	Dwie ważne równości: $p(y, \theta) = p(\theta y)p(y)$ $= p(y \theta)p(\theta)$
$p(\theta)$	Rozkład <i>a priori</i> – jest <i>brzegowym rozkładem parametrów</i> ; odzwierciedla wstępną (tj. przed wglądem w dane, przed zaobserwowaniem y) wiedzę badacza; tym samym stanowi	Prawdziwa (choć, szczerze mówiąc, mało pożyteczna:)) jest tożsamość:

¹ To jest swego rodzaju nadużycie, ponieważ gęstość rozkładu to nie to samo, co sam rozkład. Ponadto, pewne relacje, które zdefiniowane są dla *rozkładów* prawdopodobieństwa nie muszą zachodzić dla ich *gęstości* – przykładem jest równość rozkładów *ciągłych* zmiennych losowych, która wcale nie wymaga równości *wszędzie* (tj. w całej dziedzinie) ich funkcji gęstości – wymaga „jedynie” równości *prawie wszędzie* (tj. z dokładnością do zbiorów miary Lebesgue’a zero).

	<p>formalną reprezentację jego niepewności co do parametrów.</p> <p>Rozkład <i>a priori</i> dobierany jest arbitralnie (według uznania badacza).</p> <p>Parametry rozkładu <i>a priori</i> nazywamy <i>hiperparametrami</i>.</p> <p>Przykładowo, jeśli przyjmiemy, że θ jest skalarną zmienną losową ($\theta \in \mathbb{R}$) i $p(\theta) = f_N(\theta \mu, \sigma^2)$, to μ i σ^2 są <i>hiperparametrami</i>.</p>	$p(\theta) = \int_Y p(y, \theta) dy$ $= \int_Y p(\theta y) p(y) dy$
<p>$p(\theta y)$</p>	<p>Rozkład <i>a posteriori</i> – jest <i>warunkowym względem obserwacji rozkładem parametrów</i>; odzwierciedla końcową (wynikową, tj. po wglądzie w dane, po zaobserwowaniu y) wiedzę badacza o parametrach, stanowi formalną reprezentację niepewności związanej z parametrami po uwzględnieniu informacji zawartej w konkretnych danych (w danym, już nielosowym, wektorze y).</p>	<p>Kluczowa we wnioskowaniu bayesowskim jest tożsamość wynikająca z wzoru Bayesa:</p> $p(\theta y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y \theta) p(\theta)}{p(y)}$ <p>W $p(\theta y)$ argumentem jest θ, natomiast y traktujemy jako ustalony wektor obserwacji (punkt w przestrzeni obserwacji). Dlatego w powyższym zapisie $p(y)$ – niebędące funkcją θ – traktowane jest jako pewna stała. Stąd możemy zapisać poniższą <i>równość z dokładnością do stałej (mnożnikowej)</i>:</p> $p(\theta y) \propto p(y, \theta).$ <p>Dodajmy jeszcze, że ów $p(y)$ jest <i>stałą normującą</i>, która zapewnia, że $p(\theta y)$ – jako gęstość pewnego rozkładu prawdopodobieństwa – całkuje się do jedności:</p> $\int_{\Theta} p(\theta y) d\theta = \int_{\Theta} \frac{p(y, \theta)}{p(y)} d\theta =$ $= \frac{1}{p(y)} \int_{\Theta} p(y, \theta) d\theta = \frac{1}{p(y)} p(y) = 1$
<p>$p(y \theta)$</p>	<p>Model próbkowy (inaczej: próbkowy rozkład obserwacji; w skrócie: rozkład próbkowy) – jest <i>warunkowym przy ustalonym wektorze parametrów rozkładem obserwacji</i>.</p> <p>Dobór postaci rozkładu próbkowego – podobnie jak rozkładu <i>a priori</i> – jest kwestią arbitralnej decyzji/założeń badacza.</p> <p>Przykładowo, badacz zakłada, że obserwacje y_1, y_2, \dots, y_T są niezależne i pochodzą z rozkładu normalnego o wartości oczekiwanej μ i wariancji σ^2. Wówczas rozkładem próbkowym jest stosowny T-wymiarowy rozkład normalny:</p> $p(y \theta) = f_N^T(y \mu \mathbf{1}_T, \sigma^2 \mathbf{I}_T),$ <p>gdzie $\mathbf{1}_T = [1 \ 1 \ \dots \ 1]'$ jest T-elementową kolumną jedynek.</p> <p>Zasadniczo, oznaczenie $p(y \theta)$ sugeruje, że argumentem jest wektor y, natomiast wektor θ jest ustalony. Tymczasem, $p(y \theta)$ często traktować będziemy jako funkcję wektora θ przy ustalonym wektorze y, co oznacza utożsamianie $p(y \theta)$ z <i>funkcją wiarygodności</i>, $L_y(\theta)$ (oznaczanej również $L(\theta y)$ lub $L(\theta; y)$). Istotnie, <i>ex definitione</i> funkcja wiarygodności przyporządkowuje</p>	<ul style="list-style-type: none"> Wobec równości: $L_y(\theta) = p(y \theta),$ wzór Bayesa możemy zapisać w postaci: $p(\theta y) = \frac{L_y(\theta) p(\theta)}{p(y)} \propto L_y(\theta) p(\theta).$ Funkcja wiarygodności (model próbkowy) pełni bardzo ważną rolę – to właśnie przez nią dane modyfikują wstępną (przed wglądem w dane) wiedzę badacza o parametrach; $L_y(\theta)$ reprezentuje informację o θ pochodzącą z danych.

wektorowi parametrów wartość funkcji gęstości rozkładu próbkowego (przy ustalonym wektorze obserwacji):

$$L_y : \Theta \ni \theta \rightarrow p(y | \theta) \in \mathbb{R}.$$

Dlatego, co do postaci analitycznej, funkcja wiarygodności i gęstość rozkładu próbkowego są równe, tzn. $L_y(\theta) = p(y | \theta)$; natomiast różnica tkwi w „optyce”:

- w $p(y | \theta)$ argumentem jest y , zaś θ jest ustalonym punktem w przestrzeni parametrów;

- w $L_y(\theta)$ argumentem jest θ , zaś y jest ustalonym punktem w przestrzeni obserwacji.

Wobec powyższego rozróżnienia warto mieć świadomość tego, że o ile $p(y | \theta)$ jest gęstością pewnego rozkładu prawdopodobieństwa (rozkładu wektora obserwacji przy ustalonych parametrach), tzn. całka po całej przestrzeni Y z $p(y | \theta)$ równa jest jedności:

$$\int_Y p(y | \theta) dy = 1,$$

o tyle funkcja wiarygodności *nie jest gęstością* żadnego rozkładu (a dokładniej: rozkładu parametrów, które są jej argumentem). Choć z jednej strony funkcja wiarygodności, na mocy konstrukcji, przyjmuje wartości w przedziale $[0, +\infty)$ – jak na każdą dobrze zdefiniowaną funkcję gęstości przystało – tak jednak, w ogólności

$$\int_{\Theta} L_y(\theta) d\theta \neq 1.$$

Wobec powyższego, w niektórych opracowaniach rozważa się także tzw. *standaryzowaną funkcję wiarygodności*:

$$\tilde{L}_y(\theta) = \frac{L_y(\theta)}{\int_{\Theta} L_y(\theta) d\theta},$$

która może już być utożsamiana z „jakąś” funkcją gęstości wektora θ , gdyż

$$\int_{\Theta} \tilde{L}_y(\theta) d\theta = \frac{1}{\int_{\Theta} L_y(\theta) d\theta} \int_{\Theta} L_y(\theta) d\theta = 1.$$

Zaznaczmy tutaj, że rozważanie standaryzowanej funkcji wiarygodności nie ma raczej znaczenia praktycznego. Z reguły bada się ją przy okazji prezentacji prostych ćwiczeń dydaktycznych, mających na celu pokazanie, że przy rosnącej liczbie obserwacji gęstość rozkładu *a posteriori*, $p(\theta | y)$, „przechodzi” od gęstości rozkładu *a priori*, $p(\theta)$, do standaryzowanej funkcji wiarygodności (tzn. $p(\theta | y)$ staje się coraz bardziej podobna do $\tilde{L}_y(\theta)$). „Dowodzi” to faktu, że wraz ze wzrostem liczby obserwacji rośnie znaczenie informacji zawartej w danych w formowaniu się końcowej (po wglądzie w dane) wiedzy badacza o parametrach modelu statystycznego.

$p(y)$

Brzegowy rozkład obserwacji (inaczej: brzegowa gęstość obserwacji; nieco ściślej: funkcja gęstości brzegowego rozkładu

• gdy θ jest ciągłą zmienną losową:

	<p>wektora obserwacji) – reprezentuje szanse zaobserwowania wektora y, „niezależnione” od parametrów.</p> <p>Uwaga „na przyszłość”: W praktyce wartość $p(y)$ jest kluczowa przy analizie <i>mocy wyjaśniającej</i> różnych modeli bayesowskich służących opisowi tego samego zjawiska (tj. tego samego wektora y). Niestety, tylko w relatywnie prostych modelach $p(y)$ daje się łatwo obliczyć; w większości przypadków konieczne jest stosowanie zaawansowanych (i następujących pewnych trudności) metod numerycznych.</p>	$p(y) = \int_{\Theta} p(y, \theta) d\theta$ $= \int_{\Theta} p(y \theta) p(\theta) d\theta$ <ul style="list-style-type: none"> • gdy θ jest dyskretną zmienną losową (co nam się nie będzie raczej przytrafiać): $p(y) = \sum_{\theta \in \Theta} p(y, \theta)$ $= \sum_{\theta \in \Theta} p(y \theta) p(\theta)$
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2. Łączne, brzegowe i warunkowe rozkłady prawdopodobieństwa we wnioskowaniu bayesowskim

We wnioskowaniu bayesowskim nieustannie będziemy pracować z „jakimiś” rozkładami łącznymi, brzegowymi i warunkowymi. Właściwe ich rozróżnianie nie jest skomplikowane, jednak wymaga zwrócenia uwagi na fakt, iż częstokroć na dany rozkład będziemy mogli spojrzeć *jednocześnie* (choć z *różnej* perspektywy) jako rozkład łączny, brzegowy i warunkowy. Rozważmy następujący...

PRZYKŁAD. Niech wektor θ składa się z trzech współrzędnych – skalarnych parametrów θ_1, θ_2 i θ_3 :

$$\theta = [\theta_1 \ \theta_2 \ \theta_3]' \in \mathbb{R}^3.$$

W kontekście **rozkładu a priori**, $p(\theta)$:

- $p(\theta_i)$, dla dowolnego $i \in \{1, 2, 3\}$, to oczywiście *brzegowy rozkład a priori* i -tego parametru;
- $p(\theta)$ to *łączny rozkład a priori* wszystkich trzech parametrów. Jednakże, wobec tożsamości:

$$p(\theta) = \int_Y p(y, \theta) dy,$$

na $p(\theta)$ można popatrzeć jako *brzegowy* (względem obserwacji, y) rozkład parametrów modelu. Przy okazji zauważmy, że zabieg „ubrzegowienia” dokonywany jest poprzez „wyciąłkowanie” „niechcianej” zmiennej (tu: y) ze stosownego rozkładu łącznego (tu: $p(y, \theta)$) lub – wyrażając się nieco precyzyjniej – scałkowanie funkcji gęstości rozkładu łącznego względem odpowiedniej zmiennej (tu: y);

- jeśli rozważany rozkład *a priori* miałby strukturę „z zależnościami” (inaczej: „z brakiem niezależności”) pomiędzy parametrami θ_1, θ_2 i θ_3 , np. postaci

$$p(\theta) = p(\theta_3 | \theta_2, \theta_1) p(\theta_2 | \theta_1) p(\theta_1), \tag{1}$$

to dla parametru θ_1 mamy rozkład *brzegowy*: $p(\theta_1)$; dla parametru θ_2 mamy rozkład *warunkowy* względem θ_1 : $p(\theta_2 | \theta_1)$; w końcu, dla θ_3 mamy także rozkład *warunkowy*: $p(\theta_3 | \theta_2, \theta_1)$. Przy okazji, zauważmy, że iloczyn dwóch ostatnich czynników w wyrażeniu (1) daje nam rozkład *a priori* parametrów θ_1 i θ_2 :

$$p(\theta_2 | \theta_1) p(\theta_1) = p(\theta_1, \theta_2).$$

Ten ostatni jest, oczywiście, rozkładem *łącznym* dla parametrów θ_1 i θ_2 . Jednak z drugiej strony jest zarazem rozkładem *brzegowym* względem parametru θ_3 , który to zostaje „wyciąłkowany” po drodze:

$$\begin{aligned}
p(\theta_1, \theta_2) &= \int_{\mathbb{R}} p(\theta_1, \theta_2, \theta_3) d\theta_3 = \int_{\mathbb{R}} p(\theta_3 | \theta_2, \theta_1) p(\theta_2 | \theta_1) p(\theta_1) d\theta_3 = \\
&= p(\theta_2 | \theta_1) p(\theta_1) \int_{\mathbb{R}} p(\theta_3 | \theta_2, \theta_1) d\theta_3 = p(\theta_2 | \theta_1) p(\theta_1) \cdot 1 = p(\theta_1, \theta_2)
\end{aligned}$$

- jeśli rozważany rozkład *a priori* miałby strukturę „z niezależnością” (innymi słowy: o parametrach θ_i , $i = 1, 2, 3$, zakładalibyśmy, że są *a priori* niezależne), to rozkład *łączy* daje się sfaktoryzować jako iloczyn rozkładów *brzegowych*:

$$p(\theta) = \prod_{i=1}^3 p(\theta_i).$$

W kontekście **rozkładu *a posteriori***, $p(\theta | y)$:

- $p(\theta | y)$ to łączny rozkład *a posteriori* wszystkich parametrów; *ex definitione* jest rozkładem warunkowym względem wektora obserwacji;
- może się zdarzyć, że interesuje nas tylko rozkład $p(\theta_1, \theta_2 | y)$ (bo θ_3 jest np. parametrem zakłócającym, albo z innych przyczyn) – rozkład ten jest:

- *warunkowy* względem obserwacji (wszak jest ci on rozkładem *a posteriori*),
- *brzegowy* względem parametru θ_3 :

$$p(\theta_1, \theta_2 | y) = \int_{\mathbb{R}} p(\theta_1, \theta_2, \theta_3 | y) d\theta_3,$$

- *łączy* dla parametrów θ_1 i θ_2 .

[Koniec przykładu]

Na zakończenie tej sekcji dodajmy jeszcze, że *brzegowy* rozkład *a posteriori* danego parametru (czy, ogólniej, danej grupy parametrów), np. $p(\theta_i | y)$, odzwierciedla – owszem – niepewność związaną z tym parametrem (parametrami), ale – co ważne – *uwzględnia przy tym niepewność związaną z wszystkimi pozostałymi parametrami*. Dzieje się tak dlatego, że rozkład $p(\theta_i | y)$ jest wynikiem scałkowania *łącznego* rozkładu *a posteriori*, $p(\theta, y)$, względem wszystkich pozostałych (w stosunku do θ_i) parametrów. Niech $\theta_{\setminus i}$ oznacza wektor otrzymany po usunięciu i -tej współrzędnej w wektorze θ , wobec czego $\theta_{\setminus i} \subseteq \mathbb{R}^{s-1}$. Wówczas $p(\theta_i | y)$ możemy zapisać w postaci $(s-1)$ -krotnej całki z gęstości łącznego rozkładu *a posteriori* względem $\theta_{\setminus i}$:

$$p(\theta_i | y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} p(\theta_1, \theta_2, \dots, \theta_s | y) d\theta_1 d\theta_2 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_s$$

lub w skrócie:

$$p(\theta_i | y) = \int_{\mathbb{R}^{s-1}} p(\theta_i, \theta_{\setminus i} | y) d\theta_{\setminus i}.$$

Dalej, ostatecznie wyrażenie możemy zapisać w postaci:

$$p(\theta_i | y) = \int_{\mathbb{R}^{s-1}} p(\theta_i | \theta_{\setminus i}, y) p(\theta_{\setminus i} | y) d\theta_{\setminus i},$$

na podstawie której stwierdzamy, że rozkład $p(\theta_i|y)$ jest tzw. ciągłą mieszkanką warunkowych rozkładów *a posteriori*, tj. $p(\theta_i|\theta_{-i},y)$, z gęstością mieszającą $p(\theta_{-i}|y)$.²

² „A skądże mi to...?”, mógłby ktoś zapytać. Ciągłe mieszanki rozkładów prawdopodobieństwa (ang. *continuous mixtures of distributions*) są uogólnieniem mieszanek dyskretnych (ang. *discrete mixtures of distributions*). Dla przykładu rozważmy tu dyskretną mieszkankę dwóch (niezależnych) 1-wymiarowych rozkładów normalnych o gęstościach, odpowiednio, $f_{N,1}(\cdot|\mu_1, \sigma_1^2)$ i $f_{N,2}(\cdot|\mu_2, \sigma_2^2)$. Rozkłady te stanowią tzw. *składniki* (inaczej: *komponenty*) mieszanki (ang. *mixture's components*). Niech $S \in \{1, 2\}$ oznacza dyskretną, 2-punktową zmienną losową, której wartości oznaczają numery komponentów. Rozkład zmiennej losowej S jest tzw. *rozkładem mieszającym*. Dalej, niech x oznacza zmienną losową będącą ową (dyskretną) mieszkanką rozkładów normalnych. Przy okazji wyjaśnijmy, że mieszanki dyskretnie nazywamy właśnie dyskretnymi z tego względu, że rozkład prawdopodobieństwa, który miesza składnikami mieszanki, jest rozkładem skokowym (dyskretnym) – o masie prawdopodobieństwa skupionej w tylu punktach, ile mamy składników mieszanki. W prezentowanym tu przykładzie przyjmijmy, że $\Pr(S=1) = 0,6$, wobec czego $\Pr(S=2) = 1 - \Pr(S=1) = 0,4$. Gęstość rozkładu rozważanej mieszanki możemy zapisać jako

$$p(x) = f_{N,1}(x|\mu_1, \sigma_1^2) \cdot \Pr(S=1) + f_{N,2}(x|\mu_2, \sigma_2^2) \cdot \Pr(S=2) = \sum_{k=1}^2 f_{N,k}(x|\mu_k, \sigma_k^2) \cdot \Pr(S=k).$$

Dalej, zauważmy, że rozkłady $f_{N,k}(x|\mu_k, \sigma_k^2)$ ($k=1, 2$) są niczym innym jak rozkładami warunkowymi zmiennej losowej x względem S , tzn.

$$p(x|S=1) = f_{N,1}(x|\mu_1, \sigma_1^2), \quad p(x|S=2) = f_{N,2}(x|\mu_2, \sigma_2^2).$$

Z kolei prawdopodobieństwa $\Pr(S=k)$ ($k=1, 2$) stanowią brzegowy rozkład zmiennej losowej S (i mogą być postrzegane jako „wagi” komponentów mieszanki). Dlatego też iloczyn $f_{N,k}(x|\mu_k, \sigma_k^2) \cdot \Pr(S=k)$ stanowi nic innego jak łączny rozkład zmiennych losowych x i S . Zatem

$$p(x) = \sum_{k=1}^2 f_{N,k}(x|\mu_k, \sigma_k^2) \cdot \Pr(S=k) = \sum_{k=1}^2 p(x, S=k),$$

co oznacza, że brzegowa gęstość zmiennej losowej x (czyli $p(x)$) jest wynikiem „wyciąkowania” zmiennej S z rozkładu łącznego, $p(x, S)$. Podsumowując, brzegowy rozkład zmiennej losowej x , czyli $p(x)$, jest dyskretną mieszkanką (dwóch) rozkładów normalnych z rozkładem mieszającym $\Pr(S=k)$ ($k=1, 2$).

Tymczasem, w mieszankach *ciągłych* komponenty mieszanki mieszane są względem pewnego *ciągłego* rozkładu prawdopodobieństwa. Wówczas operator sumowania zastępujemy stosowną całką. W zasadzie, zawsze wtedy, gdy mamy do czynienia z jakimś rozkładem brzegowym, możemy patrzeć nań właśnie jak na mieszkankę odpowiednich rozkładów warunkowych z odpowiednią gęstością mieszającą. Dla przykładu rozważmy model z dwoma parametrami, tj. $\theta = [\theta_1 \ \theta_2]' \in \mathbb{R}^2$, w dalszym ciągu skupiając uwagę na brzegowym rozkładzie θ_1 .

- W kontekście rozkładu *a priori* mamy:

$$p(\theta_1) = \int_{\mathbb{R}} p(\theta_1, \theta_2) d\theta_2 = \int_{\mathbb{R}} p(\theta_1|\theta_2) p(\theta_2) d\theta_2.$$

Rozkład $p(\theta_1)$ stanowi mieszkankę warunkowych rozkładów *a priori* $p(\theta_1|\theta_2)$ (używam liczby mnogiej: *rozkładów*, gdyż dla różnych wartości θ_2 mamy *różne* – choć pochodzące z tej samej rodziny – rozkłady $p(\theta_1|\theta_2)$). Gęstością mieszającą jest tu brzegowy rozkład *a priori* parametru θ_2 , tj. $p(\theta_2)$.

- W kontekście rozkładu *a posteriori*:

$$p(\theta_1|y) = \int_{\mathbb{R}} p(\theta_1, \theta_2|y) d\theta_2 = \int_{\mathbb{R}} p(\theta_1|\theta_2, y) p(\theta_2|y) d\theta_2.$$

Rozkład $p(\theta_1|y)$ stanowi mieszkankę warunkowych rozkładów *a posteriori* $p(\theta_1|\theta_2, y)$ (znowu, liczba mnoga: *rozkładów!*), natomiast gęstością mieszającą jest brzegowy rozkład *a posteriori* parametru θ_2 , tj. $p(\theta_2|y)$.

[Ł.K.: Przypisy tej długości powinny być zabronione... ;)]